La lettre innovation et prospective de la CNIL

N°08 / novembre 2014

IP INNOVATION & PROSPECTIVE

Retrouvez-nous sur notre site [www.cnil.fr/ip] en flashant le code ou sur :

Mobilitics, saison 2 : Les smartphones et leurs apps sous le microscope de la CNIL et d'Inria

La CNIL et Inria travaillent depuis maintenant 3 ans sur un projet de recherche et d'innovation ambitieux nommé Mobilitics. Son objectif : mieux connaître les smartphones, ces objets utilisés quotidiennement par des dizaines de millions de français et qui restent de véritables boîtes noires pour les utilisateurs, les chercheurs et les autorités de régulation. Pourtant, ces « amis qui nous veulent du bien » sont d'extraordinaires producteurs et consommateurs de données personnelles. Du point de vue de la recherche, ils incarnent idéalement les enjeux au cœur de l'activité de l'équipe Privatics d'Inria : comprendre les mécanismes techniques autour des données personnelles et concevoir des solutions techniques préservant la vie privée. Un outil capable de détecter les accès à des données personnelles dans les appareils (localisation, photos, carnet d'adresses) a donc été développé, mis au point et expérimenté. Après une première vague de tests en 2013, une « deuxième saison » de Mobilitics a eu lieu pendant l'été 2014. Les premiers résultats présentés dans cette lettre illustrent bien l'intérêt du partenariat entre Inria et la CNIL : des outils imaginés et conçus ensemble sont utilisés par les deux institutions, chacune dans son rôle. Pour la CNIL, il s'agit de mieux comprendre ce qui se passe réellement lors de l'usage de ces appareils, pour définir des priorités d'action et émettre des recommandations. Pour Inria, il s'agit aussi de pousser plus loin les investigations et analyses techniques et de développer des solutions permettant de mieux protéger les utilisateurs. Ces travaux sont donc l'occasion pour les deux institutions de partager leurs analyses et interrogations. En effet, si ces technologies offrent des services extraordinaires aux individus et sont bénéfiques pour la société, elles ne peuvent se développer que dans le respect de la vie privée et des libertés individuelles. Rendre la technologie plus transparente et plus compréhensible aux citoyens est un défi commun pour la recherche et pour l'autorité de régulation.

Claude Castelluccia,
Directeur de recherche, responsable de l'équipe Inria Privatics ■

Vincent Roca,
Chargé de recherche, membre de l'équipe Inria Privatics ■

Les données personnelles, ingrédient de base des recettes à succès sur smartphone

Dès 2011, dans son sondage «*Smartphones et données personnelles*», la CNIL avait souligné le recours massif des utilisateurs de smartphones aux applications téléchargeables. Depuis, la tendance n'a fait que s'accentuer : diverses études indiquent que les mobinautes et tablonautes français ont en moyenne une trentaine d'applications sur leur appareil[1].

L'écosystème de ces applications a de multiples facettes et modèles économiques : les applications peuvent être liées à un autre service (l'app de votre banque par exemple), financées par l'intégration de publicités, à travers des achats «*in app*» (c'est le modèle de nombreux jeux à succès) ou encore... fondées sur la collecte et la monétisation de données à des fins publicitaires par des tiers.

En réalité, de ces modèles complémentaires, le dernier est omniprésent et incarne cette économie cachée des données personnelles sur les smartphones que nous cherchons à décrypter.

Cette diversité des modèles économiques et des stratégies est difficile à appréhender, la CNIL a donc souhaité se doter d'un outil d'analyse. Développé avec Inria, Mobilitics a été installé sur des smartphones que des agents de la CNIL ont accepté d'utiliser à la place de leur téléphone personnel pendant 3 mois. Ils ont testé 189 apps iOS et 121 apps Android (voir tableau ci-dessous).

Bien sûr, il ne s'agit pas d'une étude statistique : une poignée d'utilisateurs n'est pas représentative de 30 millions de mobinautes français, et quelques centaines d'applications ne sont pas un échantillon suffisant de l'ensemble des apps disponibles au téléchargement (plus d'un million sur chacun des deux grands magasins, App Store d'Apple et Play Store d'Android). Cependant, plusieurs semaines de données permettent de comprendre de manière détaillée le fonctionnement des applications en conditions réelles.

Dès la première vague de tests sous iOS, il a été possible de tirer *trois séries d'enseignements* : d'abord, le statut particulier de la géolocalisation, reine des données du smartphone ; ensuite, la tendance des développeurs et éditeurs d'applications à recourir à des stratégies d'identification aux objectifs très divers (mesures d'audience, statistiques d'utilisation, analytics, monétisation et publicité...) ; enfin, la difficulté à corréler les accès aux données avec des actions de l'utilisateur ou des besoins légitimes des applications. Suite à ces premiers résultats, une deuxième campagne de tests a été lancée au cours de l'été 2014. D'autres volontaires ont utilisé une nouvelle version de Mobilitics, cette fois sur des téléphones Android. Ce travail a confirmé la nécessité pour la CNIL de rester vigilante vis-à-vis de l'information des utilisateurs et des outils de maîtrise des données personnelles mis à disposition par les systèmes d'exploitation et les éditeurs d'applications.

Résultats généraux, comparaison entre les deux saisons

	iOS 5 (tests de novembre 2012 à janvier 2013)		Android « Jelly Bean » (tests de juin à septembre 2014)	
Nombre d'applications	**total : 189**		**total : 121**	
Qui communiquent sur le réseau	176	93%	80	66%
Qui accèdent à l'UDID/android ID	87	46%	41	34%
Qui accèdent à la géolocalisation	58	31%	29	24%
Qui accèdent au carnet d'adresses	15	8%	20	17%
Qui accèdent au calendrier	3	2%	4	3%
Qui accèdent au nom de l'appareil	30	16%	non mesuré	
Qui accèdent au nom d'opérateur	non mesuré		28	23%
Qui accèdent à l'IMEI (identité d'équipement mobile)	non mesuré		24	20%
Qui accèdent à l'adresse MAC WiFi	non mesuré		9	7%
Qui accèdent au numéro de téléphone	non mesuré		7	6%
Qui accèdent à l'identifiant de carte SIM (ICCID)	non mesuré		6	5%
Qui accèdent à la liste des points d'accès WiFi (SSID)	non mesuré		5	4%

[1] Etudes WebObservatoire et Home Devices 1er et 2e trimestre 2014, Médiamétrie, « baromètre trimestriel du Marketing Mobile en France », Mobile Marketing Association France, et Institut Statista

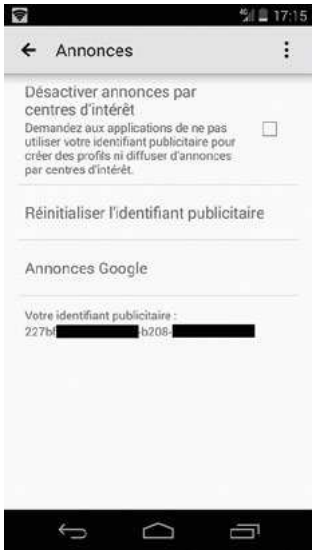
La course aux identifiants

Lors de la première vague, près d'une application testée sur deux avait accédé à un identifiant unique alphanumérique de l'appareil appelé UDID *(Unique Device Identifier)*. 40 % des applications ayant accédé à l'UDID l'avaient en outre envoyé « en clair » (c'est-à-dire sans le chiffrer) sur le réseau.

À l'époque, cet identifiant était encore accessible à toutes les applications. Les versions suivantes du système d'exploitation iOS d'Apple ont limité l'accès à cette donnée tout en créant d'autres identifiants, notamment l'*Advertising identifier*, identifiant dédié à la publicité, et l'*Identifier for vendor*, qui est unique pour chaque éditeur d'application.

Notre expérimentation Android montre des logiques similaires, mais dans un environnement par nature plus ouvert et moins contrôlé *a priori* que celui d'iOS. L'Android ID est ainsi un identifiant persistant comparable à l'UDID d'Apple tel qu'il fonctionnait en 2013, c'est-à-dire qu'il s'agit d'une série alphanumérique non modifiable. L'*advertising ID* (ad ID) est, quant à lui, un identifiant dédié à la publicité que les utilisateurs peuvent réinitialiser. Depuis le 1er août 2014, Google impose à toute nouvelle application (ou à toute application mise à jour) de n'utiliser que cet identifiant pour des fins publicitaires. En outre, les utilisateurs peuvent utiliser des réglages limitant le suivi publicitaire. Malheureusement ceux-ci ne sont ni simples à trouver, ni faciles à comprendre :

- sur les iPhones, il faut se rendre dans les réglages de confidentialité, accéder à l'option « publicité » et activer le suivi publicitaire limité ;
- sur Android, alors que l'on pourrait s'attendre à pouvoir accéder à ce paramètre dans les réglages du téléphone, il faut en réalité se rendre dans l'application « Paramètres Google » et activer une option « désactiver annonces par centres d'intérêt » (voir illustration ci-contre).



Capture d'écran de la page « annonces » de l'application « Paramètres Google » sur Android.

▶▶▶

3 questions à… Camille Gruhier

Journaliste Nouvelles Technologies, Que Choisir

Vous avez publié dans le numéro d'octobre 2014 du magazine *Que choisir* une enquête sur les applications mobiles et les données personnelles - Pourquoi votre magazine s'est-il intéressé à ce sujet ?

Que Choisir est très sensible à la problématique de la protection des données personnelles. Les applications mobiles sont désormais massivement utilisées par le grand public, chaque possesseur de smartphone en installe une trentaine en moyenne. Il était donc logique que nous analysions les échanges de données entre l'utilisateur et l'éditeur.

Quels sont les principaux enseignements que vous tirez de vos tests et qu'est-ce qui vous surprend le plus du point de vue de la protection des consommateurs ?

Nous avons constaté que plusieurs applications, même très célèbres, ne se gênent pas pour collecter des informations dont elles n'ont absolument pas besoin pour fonctionner. Cela confirme un besoin de collecter un maximum de données pour vendre aux annonceurs des fichiers très ciblés. Le plus choquant est sans doute que l'utilisateur n'est pas informé de cette collecte, on lui en demande encore moins l'autorisation.

Avez-vous donné des recommandations à vos lecteurs ?

Nous conseillons de ne pas télécharger d'applications inutiles et de faire régulièrement le tri dans son smartphone. Par ailleurs, mieux vaut éviter de se connecter depuis les réseaux WiFi publics car certains flux de données ne sont pas chiffrés.

▸▸▸

Si les traductions techniques de ces réglages sont différentes, nos tests montrent toutefois que la coexistence de nombreux identifiants facilite grandement les possibilités de contournement de ce type de cloisonnement. Tant que les pratiques de collectes multiples persistent, aucune protection « *privacy by design* » ne peut s'avérer techniquement efficace et, en dehors des garanties juridiques, seule la bonne volonté des développeurs permet de garantir qu'ils ne contournent pas les réglages mis en œuvre par les utilisateurs pour limiter le ciblage publicitaire.

En effet, les applications testées utilisent une grande diversité d'identifiants qui pourraient servir à enrichir leurs moyens de suivi de l'utilisateur. Sur le téléphone d'un utilisateur qui avait installé 58 applications, cela se traduit de la manière suivante :

- 23 applications ont accédé à l'Android_id ;
- 18 applications ont accédé à l'IMEI (*International Mobile Equipment Identity*), qui est un identifiant alphanumérique unique du téléphone permettant notamment de bloquer un terminal perdu ou volé en l'empêchant de se connecter aux réseaux des opérateurs ;
- 10 applications ont accédé à l'IMSI (International Mobile Subscriber Identity), qui est un identifiant unique stocké dans la carte SIM permettant à un opérateur d'identifier un abonné ;
- 5 applications ont eu accès à l'identifiant unique de la carte wifi (l'adresse mac) du téléphone. Cet identifiant est utilisé lors de toutes les communications entre un terminal et un point d'accès wifi. Lorsque le wifi est activé, l'adresse MAC est périodiquement diffusée par le téléphone ;
- 5 applications ont accédé tout simplement… au numéro de téléphone ;
- 5 applications ont accédé à l'identifiant unique de la carte SIM (ICCID) ;
- 2 applications ont accédé à la liste des identifiants des points d'accès wifi (Wifi_ssid) à portée de l'utilisateur.

Le smartphone, un GPS dans votre poche… pour les autres ?

Entre un quart et un tiers des applications présentes sur les différents appareils des vagues 1 (iOS) et 2 (Android) ont eu accès à la localisation de l'appareil.

Ce chiffre n'est en soi ni choquant ni surprenant : de nombreuses applications ont des raisons légitimes d'accéder à la position géographique de l'appareil, que ce soit par le GPS ou grâce à la détection des antennes du réseau cellulaire ou des bornes wifi entourant l'appareil. Sur l'ensemble des applications testées, la très grande majorité avait au moins une fonction accessoire utilisant la géolocalisation. Le plus surprenant est surtout l'intensité et la fréquence d'accès à cette information par certaines applications. Par exemple, sur une période de 3 mois, une application a accédé plus de 1 million de fois à la géolocalisation et une deuxième application plus de 700 000 fois. Cela représente en moyenne près d'un accès par minute sur une période de 3 mois… Et pourtant il ne s'agissait pas d'applications de navigation ou d'itinéraire.

La géolocalisation est donc, en volume, la donnée la plus collectée : elle représente à elle seule plus de 30% des évènements détectés par nos outils. Pour autant, rien ne permet d'affirmer que les éditeurs récupèrent en permanence ou périodiquement («par paquets») cette information : elle peut n'être collectée que pour être utilisée dans l'appareil par l'application. Ces accès si nombreux ne peuvent être reliés simplement à des fonctionnalités offertes par l'application et encore moins à une action demandée par l'utilisateur. Ils soulèvent dès lors en eux-mêmes une question de protection de la vie privée, transformant le téléphone en un instrument permanent de localisation de son propriétaire (sans même parler des conséquences éventuelles en termes de performance ou de durée de vie de la batterie).

Par exemple, si une application de réseau social peut disposer de la géolocalisation au moment où l'utilisateur souhaite partager un contenu localisé, l'accès quasi-permanent à cette information sur une longue période semble disproportionné et constitue une source de risques pour la personne géolocalisée[2].

S'agit-il d'un problème de mauvaise optimisation des commandes de l'application (qui n'auraient alors pas été pensées pour réduire la collecte de cette donnée à ce qui est utile, mais la rendraient au contraire permanente « au cas où »), ou bien certaines applications chercheraient-elles à acquérir des informations riches sur l'ensemble des localisations d'une personne, en dehors de tout lien avec les fonctionnalités premières de l'application ?

Quoiqu'il en soit, quand on connait la richesse de l'information de géolocalisation pour les objectifs de marketing, de ciblage, de contextualisation et de personnalisation, on est en droit de s'interroger sur la nécessité d'améliorer les solutions proposées actuellement par les systèmes d'exploitation des smartphones. Ces solutions reposent en effet sur une autorisation ou un refus («à prendre ou à laisser») plus ou moins générique : global pour Android ou application par application sur iOS. L'intérêt d'une amélioration en ce domaine serait important puisque les travaux de *Sébastien Gambs*, de l'IRISA, ou d'*Yves-Alexandre de Montjoye*, du MIT et de l'Université catholique de Louvain, ont notamment montré qu'une base de données de localisation permettait de déduire des informations détaillées sur les habitudes et modes de vie des personnes : lieux de vie et de travail, sorties, loisirs, mobilités, mais aussi éventuellement fréquentation d'établissements de soins ou de lieux de culte…

[2] Sur iOS8, déployé en septembre 2014, les applications peuvent demander l'accès à la localisation « toujours » ou seulement « lorsque l'app est en marche », ce qui va dans le bon sens du point de vue de la granularité des réglages.

Les systèmes d'exploitation et leurs magasins d'applications, des acteurs pas comme les autres

La saison 1 de Mobilitics avait montré une présence non négligeable d'Apple dans son propre écosystème en tant que collecteur de données issues de l'iPhone[3]. Déjà dans iOS, Google était un acteur extrêmement présent à travers ses différents services (Google maps, Gmail, Google search...) mais également en tant que fournisseur de fonctionnalités (publicité, analytics...) pour des applications d'autres éditeurs. Les résultats issus de l'étude des usages des smartphones Android renforcent ce constat, en particulier pour certains services installés par défaut sur les appareils :

- L'application Play Store est ainsi l'une des plus grosses consommatrices de données, puisqu' elle a accédé en 3 mois et pour un seul utilisateur 1 300 000 fois à la localisation cellulaire, 290 000 fois au GPS, 196 000 fois au scan du wifi et plusieurs milliers de fois à quelques autres données. C'est d'autant plus remarquable que l'app Play Store est quasiment indispensable pour installer ou mettre à jour des applications et qu'elle est devenu un élément sous-jacent du système d'exploitation pour lequel il centralise et contrôle de nombreuses fonctions et mises à jour. Un article de presse spécialisée l'a ainsi qualifié en juillet 2014 de « gardien de l'écosystème Android[4] ».
- L'application-widget « Actualités et météo » a accédé 1 560 926 fois à la localisation de l'utilisateur pendant les trois mois de l'expérimentation. Cette application a aussi communiqué 341 025 fois avec internet.

Or, ces applications étant présentes par défaut sur l'appareil et ne pouvant être supprimées, l'utilisateur n'a pas pu consulter les informations collectées, qui sont généralement affichées avant le téléchargement et l'installation d'une application sur Android. La mise en œuvre de mesures d'information et de réglages spécifiques pourrait dès lors être envisagée. Il serait par exemple possible de créer des réglages dédiés, par exemple un tableau de bord (dashboard) explicitant leurs accès et transmissions de données et les raisons associées, avec des possibilités de refuser (opt out) ou d'accepter (opt in) certaines fonctions.

[3]Par exemple, Apple utilise les iPhones pour mettre à jour ses bases de données de localisation des antennes wifi, sur lesquelles reposent ses propres services de cartographie. Apple utilise également l'envoi de données vers ses propres serveurs pour des services tel que Siri, qui est basé sur l'analyse de la voix.

[4]Ulrich ROZIER, « Google Play Services, le cheval de Troie de Google », FrAndroid, 9 juillet 2014.

Transparence et maîtrise par les utilisateurs : une responsabilité partagée par tous les acteurs du marché

De nombreux documents produits par les régulateurs ou par l'industrie elle-même mettent en avant à la fois les responsabilités qui pèsent sur chacun des acteurs ainsi que les bonnes pratiques et recommandations à respecter. Ainsi, le « groupe de l'article 29 » (le groupe des CNIL européennes) a publié en 2013 *un avis sur les applications destinées aux dispositifs intelligents*. De même, le bureau du *procureur général de Californie*, la *Federal Trade Commission*, le *GSMA* (regroupant les opérateurs mobiles), l'*autorité de protection de données de Bavière* ou l'*ICO* britannique ont publié leur propre série de recommandations. Les résultats Mobilitics montrent la nécessité pour chaque catégorie d'acteurs de prendre la mesure de ses propres responsabilités et non de faire peser cette charge sur les autres (ou sur les utilisateurs).

Les grands systèmes d'exploitation et magasins d'applications ont un rôle clé et des responsabilités lourdes. Ils définissent le cadre d'action des autres acteurs en décidant ce qui est techniquement possible et ce qui ne l'est pas, les outils d'information et de maîtrise qui sont disponibles et le moment auquel il est possible d'y accéder (au téléchargement, à l'installation, par des alertes à l'écran, dans les réglages de l'appareil). Si chaque nouvelle version des systèmes d'exploitation est l'occasion de changements souvent positifs dans les outils et règles disponibles, ces acteurs doivent renforcer leur engagement au service de la protection de la vie privée des utilisateurs. Leur responsabilité est d'autant plus grande qu'ils sont dans une situation très privilégiée par rapport à l'accès à des informations collectées ou stockées sur l'appareil (voir partie précédente). Les développeurs et éditeurs d'application doivent quant à eux adopter une approche de *privacy by design* et notamment minimiser les données en s'interdisant la collecte des données qui ne sont pas liées au service rendu par l'application.

Geoffrey Delcroix,
Chargé d'études, innovation et prospective, CNIL ■

Stéphane Petitcolas,
Ingénieur au service de l'expertise technologique, CNIL ■

CNIL
Commission Nationale de l'Informatique et des Libertés

Commission Nationale de l'Informatique et des Libertés

8, rue Vivienne - CS 30223 - 75083 Paris CEDEX 02
Tél.: 01 53 73 22 22 - **Fax:** 01 53 73 22 00
publications@cnil.fr
Édition trimestrielle
Directeur de la publication: Édouard Geffray
Rédacteur en chef: Gwendal Le Grand
Conception graphique: EFIL Communication 02 47 47 03 20 - www.efil.fr
Impression: Champagnac
Crédit photos: CNIL
ISSN: 2118-9102
Dépôt légal: à publication